Chapter 9

Documents

To improve utility of an archival product, supplemental and/or ancillary reference material (documents) may be included. This material is primarily textual and should augment, reinforce, and assist in the overall understanding and/or use of the data products or software. It may include, but is not limited to:

(1) flight project documents
(2) instrument papers
(3) science articles
(4) volume information
(5) software interface specifications (SISs)
(6) software user manuals

PDS has three guidelines for deciding whether to include documents: (1) Would this information be helpful to a data user? (2) Is the material necessary? (3) Is the documentation complete? Recognizing that there are many levels of inquiry regarding data sets, PDS seeks to err on the side of completeness.

Each document must be prepared and saved in a PDS compliant format, have a PDS compliant label, and be stored in the DOCUMENT directory of an archive volume (see the Volume Organization and Naming chapter of this document).

A readable ASCII text version of each document must be included on the archival volume; documents may also be provided in certain other PDS approved formats, at the option of the data producer. 'ASCII text' includes only printable ASCII characters and the ASCII blank space; it does not include special software constructs or graphics. 'Readable' means that special markup insertions do not significantly interfere with the user's ability to read the file. Plain ASCII text, HTML, TeX, and LaTeX files-subject to the 'readability' criterion-satisfy the 'ASCII text' PDS requirement.

9.1   PDS Objects for Documents

Either of two PDS objects-TEXT or DOCUMENT-may be used for documents, subject to the constraints in the next two subsections.

9.1.1   TEXT Objects

TEXT objects are preferred for documents which stand alone and have a narrow focus-for example, describing the contents of an archive volume (AAREADME.TXT) or the contents of a

directory (DOCINFO.TXT). Each file which uses the PDS TEXT object must:

> (a) have an attached or detached label containing a TEXT object definition (see Appendix A);
>
> (b) be in plain, unmarked ASCII text (printable ASCII characters and the ASCII blank space), cannot include programming language constructs such as HTML, and cannot include graphics; and
>
> (c) have a file name which ends with the extension .TXT.

9.1.2   DOCUMENT Objects

DOCUMENT objects are preferred when several versions of the same file are provided and/or when there are several component files in the document-for example, graphics in addition to text. Each file which uses the PDS DOCUMENT object must:

> (a) have a detached label (or, in the case of a single file in plain ASCII text, optionally an attached label) containing a DOCUMENT object definition (see Appendix A);
>
> (b) be in plain, unmarked ASCII text; an approved markup language format; or an approved non-ASCII format; and
>
> (c) use the objects, interchange formats, document formats, and file name extensions below:

Format	Object	Interchange Format	Document Format	File Extension
Plain ASCII Text	ASCII_DOCUMENT	ASCII	TEXT	.ASC
HTML	HTML_DOCUMENT	ASCII	TEXT	.HTM or .HTML*
TeX	TEX_DOCUMENT	ASCII	TEXT	.TEX
LaTeX	LATEX_DOCUMENT	ASCII	TEXT	.TEX
Adobe PDF	PDF_DOCUMENT	BINARY	ADOBE PDF	.PDF
MS Word	WORD_DOCUMENT	BINARY	MICROSOFT WORD	.DOC
Rich Text	RTF_DOCUMENT	BINARY	RICH TEXT	.RTF
GIF	GIF_DOCUMENT	BINARY	GIF	.GIF
JPG	JPG_DOCUMENT	BINARY	JPG	.JPG
Encapsulated Postscript	EPS_DOCUMENT	BINARY	ENCAPSULATED POST SCRIPT	.EPS
Postscript	PS_DOCUMENT	BINARY	POSTSCRIPT	.PS
Tagged Image File Format	TIFF_DOCUMENT	BINARY	TIFF	.TIF or .TIFF*

* See chapter *File Specification and Naming* regarding extensions with more than three characters.

Example: A document defined as a DOCUMENT object appearing in the DOCUMENT directory might be the following:

    1. MYDOC.ASC   - required ASCII version
    2. MYDOC.DOC   - optional Microsoft Word version to retain all graphics
    3. MYDOC001.TIF   - optional scanned TIFF version of selected pages
    4. MYDOC002.TIF   - optional scanned TIFF version of other selected pages
    5. MYDOC.LBL   - PDS label defining DOCUMENT object(s) for these files

Optional versions of the document should have the same file name as the required ASCII version but with different extensions. Optional versions should be defined as DOCUMENT objects in the PDS label; the name of the required ASCII file should be indicated using the DESCRIPTION keyword.

9.2   Document Format Details

9.2.1   Plain, Unmarked ASCII Text

Line Delimiters / Line Length - PDS recommends that lines of plain ASCII text be limited to 78 or fewer characters. The line length recommendation is made to facilitate importation of text into environments which may reserve several characters for line numbering or other uses. Regardless of length, each line must be followed by the Carriage Return (ASCII decimal 13) and Line Feed (ASCII decimal 10) characters. The use of the Carriage Return and Line Feed characters ensures readability in the four environments commonly in use by planetary researchers. In the Macintosh and Unix environments simple utilities are available (Apple File Exchange and Translate, respectively) to add (if submitting data) or eliminate (if using data) the Line Feed or Carriage Return.

Page Break / Page Length - Paragraphs should be separated by one or more empty lines-ASCII blanks only (if used for padding) and the Carriage Return/Line Feed sequence. This facilitates simple conversion of text files into word processor formats. In order to organize text into pages, the Page Feed (ASCII decimal 12) is allowed. If this feature is used, page length should be limited to 60 lines of text and a Page Feed character should be inserted immediately after the END (Carriage Return/Line Feed) statement of any PDS labels which appear at the beginning of the file.

9.2.2   ASCII Text Containing Markup Language

Line Delimiters / Line Length - Because these files conform to specific markup language constructs, the recommended line length of 78 or fewer characters does not apply. However, each line must be delimited by the Carriage Return (ASCII decimal 13) and Line Feed (ASCII decimal 10) characters.

Page Break / Page Length - Because these files conform to specific markup language constructs and because the Page Feed (ASCII decimal 12) character may not function as such, the recommended page length limit does not apply.

Note: There are several markup languages which are not recommended because the markup is so extensive that it inhibits easy reading of the text portions of the file. There are also automatic generators or converters of files that create output in approved formats which are also difficult to read. Such files may be judged unsuitable for meeting the 'ASCII text' requirement.

The following subsections describe PDS standards applicable to HTML files.

9.2.2.1   HyperText Markup Language (HTML) Version

PDS archive products must adhere to Version 3.2 of the HTML language, a Standard Generalized Markup Language which conforms to International Standard ISO 8879. All files are subject to validation against the HTML 3.2 SGML Declaration and the HTML Document Type Definition.

Note: Constructs not defined in the HTML 3.2 standard (e.g., FRAME, STYLE, SCRIPT, and FONT FACE tags) are not allowed in PDS document files.

9.2.2.2   Location of Files

PDS strongly recommends that targets of all HTML links be present on the archive volume. In cases where external links are provided, the link should lead to supplementary information which is not essential to understanding or use of the archival data.

PDS recommends that all files which comprise an HTML document-or series of documents-be co-located in a single directory. However, locating ancillary files (e.g., images, common files) in subdirectories may be required under certain circumstances (e.g., to avoid conflicts in file names and to minimize replication of common files).

9.2.2.3   Character Set

Documentation files prepared to accompany the data set or data set collection must be validated in that the files can be copied or transmitted electronically, and can be read or printed by their target text processing program. Documentation files should be spell-checked prior to being submitted to PDS for validation.

9.2.2.4   Excluded / Discouraged HTML 3.2 Capabilities

Although the APPLET tag is advertised to be supported by all Java enabled browsers, not all applets execute on all browsers on all platforms. Further, some browsers require that the user explicitly enable use of Java applets before the applet will execute. Applets are permitted only

when the information they convey is not essential to understanding or use of the archival data.

Use of the TAB character is permitted but strongly discouraged because of variations in implementation among browsers and resulting misalignments within documents.

Use of animated GIFs is discouraged.

9.2.3   Non-ASCII Formats

Where possible the version or ENCODING_TYPE should be specified for the format used. For example, a PostScript or Encapsulated PostScript file could be further documented by noting the generating version with the DESCRIPTION keyword. The label defining a GIF document might include the keyword-value pair ENCODING_TYPE = "GIF87A".

9.3   Examples

9.3.1   Simple Attached label (Plain ASCII Text)

The following label could be prepended to a plain ASCII text file which describes the content and format of Mars Pathfinder Imager Experiment Data Records. The aggregate file could be stored under the file name EDRSIS.TXT.

```
PDS_VERSION_ID                   = PDS3
RECORD_TYPE                      = STREAM
OBJECT                           = TEXT
 NOTE                            = "Mars Pathfinder Imager Experiment Data Record SIS"
 PUBLICATION_DATE                = 1998-06-30
END_OBJECT                       = TEXT
END
```


9.3.2   Complex Detached Label (Two Document Versions)

If the data producer chose to provide the same document in both plain ASCII text and as a Microsoft Word document, the detached label would have the name EDRSIS.LBL and would be as follows:

```
PDS_VERSION_ID                   = PDS3
RECORD_TYPE                      = UNDEFINED
^ASCII_DOCUMENT                  = "EDRSIS.ASC"
^WORD_DOCUMENT                   = "EDRSIS.DOC"

OBJECT                           = ASCII_DOCUMENT
 DOCUMENT_NAME                   = "Mars Pathfinder Imager Experiment Data Record"
 PUBLICATION_DATE                = 1998-06-30
 DOCUMENT_TOPIC_TYPE             = "DATA PRODUCT SIS"
 INTERCHANGE_FORMAT              = ASCII
```

```
DOCUMENT_FORMAT              = TEXT
DESCRIPTION                  = "This document contains a textual description
                                of the VICAR and PDS formatted Mars Pathfinder
                                IMP Experiment Data Records.  This is the ASCII
                                text version of the document required by PDS."
END_OBJECT                   = ASCII_DOCUMENT

OBJECT                       = WORD_DOCUMENT
 DOCUMENT_NAME               = "Mars Pathfinder Imager Experiment Data Record"
 PUBLICATION_DATE            = 1998-06-30
 DOCUMENT_TOPIC_TYPE         = "DATA PRODUCT SIS"
 INTERCHANGE_FORMAT          = BINARY
 DOCUMENT_FORMAT             = "MICROSOFT WORD"
 DESCRIPTION                 = "This document contains a textual description
                                 of the VICAR and PDS formatted Mars Pathfinder
                                 IMP Experiment Data Records.  The document was
                                 created using MicroSoft Word 6.0.1 for the Macintosh."
END_OBJECT                   = WORD_DOCUMENT
END
```

9.3.3   Complex Detached Label (Documents Plus Graphics)

The following label (EDRSIS.LBL) illustrates the use of an HTML document as the required
ASCII document. The same document is also included as a PDF file, and four GIF images are
included separately.

```
PDS_VERSION_ID               = PDS3
RECORD_TYPE                  = UNDEFINED
^HTML_DOCUMENT               = "EDRSIS.HTM"
^PDF_DOCUMENT                = "EDRSIS.PDF"
^GIF_DOCUMENT                = ("FIG1.GIF","FIG2.GIF","TAB1.GIF","TAB2.GIF")

OBJECT                       = HTML_DOCUMENT
 DOCUMENT_NAME               = "Mars Pathfinder Imager Experiment Data Record"
 PUBLICATION_DATE            = 1998-06-30
 DOCUMENT_TOPIC_TYPE         = "DATA PRODUCT SIS"
 INTERCHANGE_FORMAT          = ASCII
 DOCUMENT_FORMAT             = HTML
 DESCRIPTION                 = "This document contains a textual description
                                 of the VICAR and PDS formatted Mars Pathfinder
                                 IMP Experiment Data Records.  This is the ASCII
                                 text version of the document required by PDS."
END_OBJECT                   = HTML_DOCUMENT

OBJECT                       = PDF_DOCUMENT
 DOCUMENT_NAME               = "Mars Pathfinder Imager Experiment Data Record"
 PUBLICATION_DATE            = 1998-06-30
 DOCUMENT_TOPIC_TYPE         = "DATA PRODUCT SIS"
 INTERCHANGE_FORMAT          = BINARY
 DOCUMENT_FORMAT             = "ADOBE PDF"
DESCRIPTION                  = "This document contains a textual description
                                of the VICAR and PDS formatted Mars Pathfinder
                                IMP Experiment Data Records.  This is the ASCII
                                text version of the document required by PDS."
END_OBJECT                   = PDF_DOCUMENT

OBJECT                       = GIF_DOCUMENT
 DOCUMENT_NAME               = "Mars Pathfinder Imager Experiment Data Record"
 PUBLICATION_DATE            = 1998-06-30
 DOCUMENT_TOPIC_TYPE         = "DATA PRODUCT SIS"
```

```
FILES                        = 4
ENCODING_TYPE                = "GIF89A"
INTERCHANGE_FORMAT           = BINARY
DOCUMENT_FORMAT              = GIF
DESCRIPTION                  = "This document is a GIF89A representation of
                                two figures and two tables from the Mars
                                Pathfinder IMP Experiment Data Record SIS.
                                The figures and tables are included in the
                                PDF version of the document; only the tables
                                are included in the HTML version -- and then
                                in a somewhat different presentation than in
                                the original, printed version of the document.
                                The files use the naming convention FIGxx.GIF,
                                where xx is the figure number, and TABxx.GIF,
                                where xx is the table number, in the file."
END_OBJECT                   = GIF_DOCUMENT
END
```